

N82-

NASA CONTRACTOR REPORT 166352

NASA-CR-166352
19820018887

Exploration into Technical Procedures
for Vertical Integration

Russell J. Michel
Keith D. Maw

Environmental System Research Institute

FOR REFERENCE

NOT TO BE TAKEN FROM THIS ROOM

LIBRARY COPY

SEP 14 1982

Purchase Order No. A-65111B
September 1979

LANGLEY RESEARCH CENTER
LONDON, ENGLAND
RESEARCH, VIRGINIA

NASA



NF02618

NASA CONTRACTOR REPORT 166352

Exploration into Technical Procedures
for Vertical Integration

Russell J. Michel
Keith D. Maw

Environmental System Research Institute
380 New York Street
Redlands, California 94035

Prepared for
Ames Research Center
under Purchase Order No. A-65111B



National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035

N82-26763#

TABLE OF CONTENTS

	<u>Page</u>
Introduction	1
Steps in the Creation of the Geo-based Information System	5
I. Definition of the Study Area	5
II. Mapping the Data	5
III. Digitizing	7
IV. Automation	8
V. Conversion to Image Format	9
Procedures in Preparation and Classification of Landsat Data	10
I. Obtain Data	10
II. Define Study Boundary	10
III. Registration to UTM Grid	11
IV. Conversion to GBIS Format	12
Integration of Data Sets	13
I. Integration of GBIS with Landsat	14
II. Integration of Landsat Classified Image with a GBIS	19
Conclusion	23

INTRODUCTION

The California Environmental Data Center is concerned with increasing the uses and benefits of the Landsat program. Landsat imagery and the information which can be obtained from Landsat can play important roles in the information systems of many public agencies, but this is not now widely recognized. While Landsat imagery has been used in a number of large scale projects, its use in smaller projects, especially those closely tied to agency's ongoing and day-to-day operations, has been much less frequent. For most such agencies, Landsat data can play only a secondary role as a data gathering system; they rely for most of their data on more conventional sources. This current inadequacy in the present system of coordination and availability does not allow this valuable data to be used to its full potential. At the federal level, NASA is concerned with the same issue.

A central theme in work undertaken by these two agencies is efficient gathering of environmental data for management of our natural resources, and the transfer of necessary technologies to appropriate levels of both government and private enterprise.

A joint effort designed to address the needs of both agencies has resulted in the initiation of The California Integrated Remote Sensing System (CIRSS) Project. This project is sponsored by NASA and involves three studies:

1. Vertical Integration
2. State User Needs
3. Landsat Demonstration Projects

Environmental Systems Research Institute (ESRI) of Redlands, California, is a research and consulting organization that centers on the analysis and interpretation of geographic information for planning and decision making.

ESRI recently participated in a pilot project with NASA Ames Research Center (NASA/ARC) which was designed to integrate Landsat imagery with an automated geo-based information system (GBIS). While ESRI currently uses Landsat photographic products as collateral information in the creation of GBIS data bases by photo interpretation, the two systems have not been operationally integrated and used together. If Landsat data is to be an effective part of public agencies and private sector information systems, Landsat data must be integrated into the existing data systems. Moreover, it must be integrated in ways which are as compatible as possible with the existing system, requiring the least change on the part of the user's existing procedures.

The Vertical Data Integration Study under CIRSS is designed to accomplish these purposes. It begins with user needs and goes from these to the data available from Landsat. It is designed to produce useful products to the agencies and to the private sector, in formats convenient for their operational use.

Vertical data integration under optimum conditions, high technology operations or operations requiring large scale bulk processing would be performed at a central location by higher organizational levels. Lower technology efforts that emphasize ancillary data and intense field work or review would be performed at lower levels. The exact mix of function and organizational level would likely vary with data sets and organizational context. This is seen as a key concept in the CIRSS project. That is, a strategy for the creation of data sets and accompanying analysis techniques that allow data sets to be meaningfully applied at various jurisdictional levels, corresponding scales, and resolutions.

It is important to recognize that vertical integration requires as much emphasis on the existing data bases as it does on the Landsat data; integration requires clear understanding of where the data are going and how to enter the data. This integration may employ either "bottom-up" or "top-down" techniques depending upon the particular data sets, their characteristics, institutional frameworks, and available processing technologies.

Examples of "bottom-up" versus "top-down" integration include:

1. Land Ownership Data

The original data collection is done by the county or city agency where the data is often used as individual records. Data flow in the upward direction would generally be accompanied by either geographical or categorical aggregation.

2. Satellite Data

The initial data collection is done through high altitude sensors employing ultra-high technology with highly centralized processing and analysis functions. The data is passed downward at the original resolution and information content. Spatial disaggregation, in a sense, might occur with increasingly accurate geometric corrections. Categorical and spatial data integrity could be achieved by modeling relationships with locally collected data of higher resolution.

This paper examines two prototypical procedures for integrating Landsat and traditional geo-based data files. In the first case, geo-based grid files were transferred to the Electromagnetic Systems Laboratory (ESL) Interactive Digital Image Manipulation System (IDIMS) at NASA/Ames. In the second case, image files from IDIMS were transferred to a traditional GBIS created and operated by ESRI. Before discussing the integration procedures used for

this pilot project, this paper will address the data sources, automation procedures, and creation of the grid cell data base at ESRI, and will also address the data sources and classification scheme of the Landsat image at NASA/ARC. The integration procedures are then discussed along with problems and solutions for this integration study. Conclusions of this pilot study are stated in the final section of this report.

STEPS IN THE CREATION OF THE GEO-BASED INFORMATION SYSTEM

I. Geographical Definition of the Study Area.

ESRI chose as a pilot study area, a portion of the Redlands, California 7.5 minute U.S.G.S. quadrangle. The study area was 7072 hectares (approximately twenty-four square miles), occupying the central 75% of the quadrangle. The north-west coordinate of the study area was located at the UTM coordinates of 479,000 E and 3,773,000 N, and an 80 meter square grid pattern was aligned with the UTM map grid. The area included the central business district, residential areas, outlying urban fringes, citrus orchards, chaparral-covered foothills, and valley bottom floodplains. For such a small geographic area, the data types per variable were quite diverse.

II. Mapping of the Data

A. Data Sources

The following data sources were used as collateral information:

1. Low altitude, false color, infrared imagery at a scale of 1:24,000.
2. Existing maps.
 - a. 1974 Land Use Maps that were prepared at ESRI for Southern California Edison.
 - b. Census tracts.
 - c. Zoning.
 - d. General Plan.
 - e. Assessor's Parcel.
 - f. U.S.G.S. Topographic.

3. Existing Reports.

- a. Soil Conservation Service report on prime agricultural lands.

B. Creating the 1978 Land Use Map

All the data variables composing the 1978 Land Use Map were reformatted to a consistent scale of 1:24,000; the imagery was used to update and verify the existing maps. A manuscript map was drafted on stable base mylar which contained all the data variables. Therefore, the manuscript map consisted of sub-areas (polygons), each with an associated code string which numerically describes each variable. The product of this cartographic effort is called an Integrated Terrain Unit Map (ITUM)¹. For geographic referencing tic marks, representing known latitude and longitude locations, were drafted on the manuscript map. There were four tics on the map, one at each corner.

1 An ITUM is generated by superimposing several sets of data and drafting a map of the polygons created. Each polygon then represents multiple data.

III. Digitizing

Once the map has been drafted and prepared, the map is digitized, which is the process of capturing the polygon configurations in a series of cartesian coordinates which preserves the spatial integrity of the mapped data. The manuscript map was taped to the digitizing machine surface and a cross-hair cursor passed over the line segments composing the polygons. On the 1978 Land Use Map, there were approximately 2,300 polygons for the entire 7.5 minute quadrangle.

IV. Automation

After the digitized file was obtained, a series of ESRI user-oriented software was used to transform the data, which has been referenced to known latitude and longitude coordinates, into an error-free file which is an automated copy of the original manuscript map.² The codes were then associated with their respective polygons, broken into distinct variables, and a uniform grid superimposed atop the polygon files for each variable.

A square grid cell of 80 meters was chosen to minimize the effects of resampling the Landsat file. This procedure of transforming each variable into a grid cell structure allows for the overlay of individual variables which represent the same geographic area. This procedure allows for the application of varied and complex modeling procedures for the analysis of natural hazards, constraints, resources, and for the spatial definition of land capabilities and suitabilities.³

2 The actual computer programs and techniques used to produce the error-free file will not be discussed in this paper.

3 It should be noted that this mapping effort and study was initially done to aid the development of a growth management study for the City of Redlands.

The following is a list of the variables in the GBIS which was created by ESRI:

- 1) 1974 Land Use
- 2) 1978 Land Use
- 3) Enumeration Districts
- 4) Status of Undeveloped Land
- 5) Zoning of Undeveloped Land
- 6) Existing Areas of the General Plan
- 7) Proposed Areas of the General Plan
- 8) Fire Hazard Areas
- 9) Flood Prone Areas
- 10) Agricultural Soil Potential
- 11) Soil Depth

(See Figures 1, 2, 3, 4 for polygon maps and grid cell maps of the 1974 and 1978 Land Use data as automated by ESRI.)

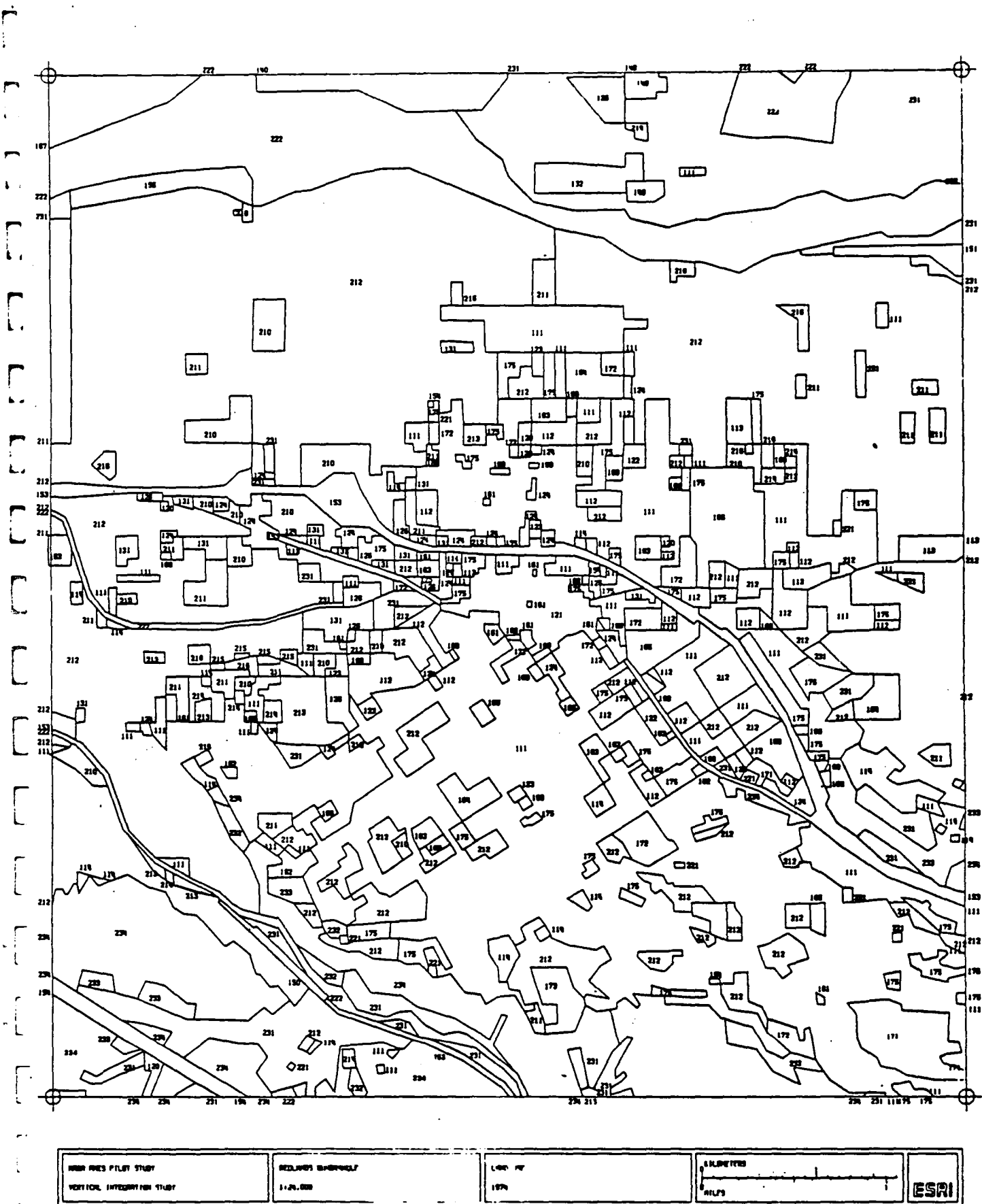
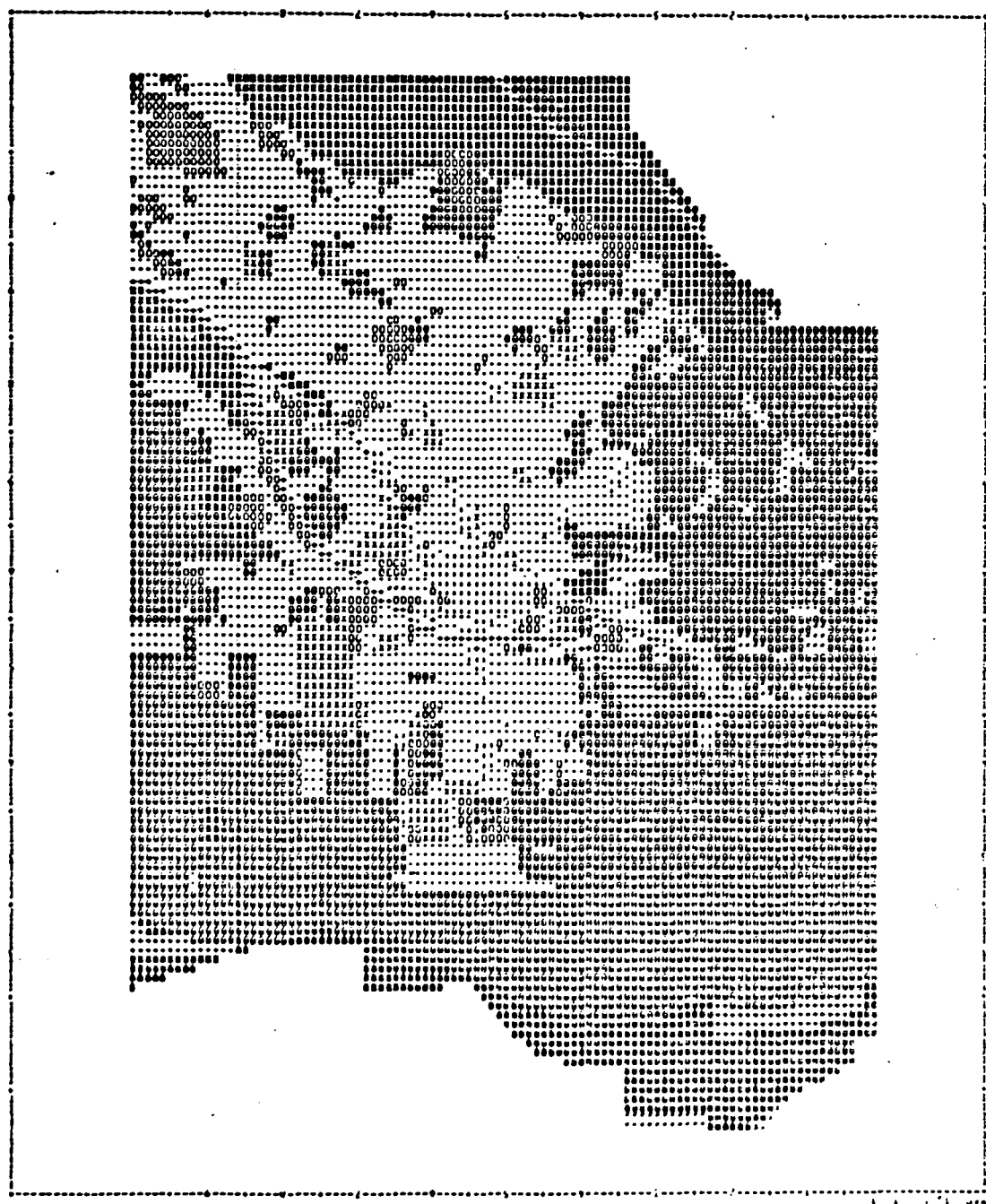


Figure 1 - 1974 Land Use

Figure 2 - 1974 Land Use



LAND USE CLASSIFICATIONS

1. AGRICULTURE AND FOREST

2. URBAN AND SUBURBAN

3. WETLANDS AND WILDLIFE

4. OPEN SPACE AND RECREATION

5. WATER BODIES

DATA SOURCES: 1. LAND USE DATA FROM 1974

2. LAND USE DATA FROM 1974

3. LAND USE DATA FROM 1974



Figure 3 - 1978 Land Use



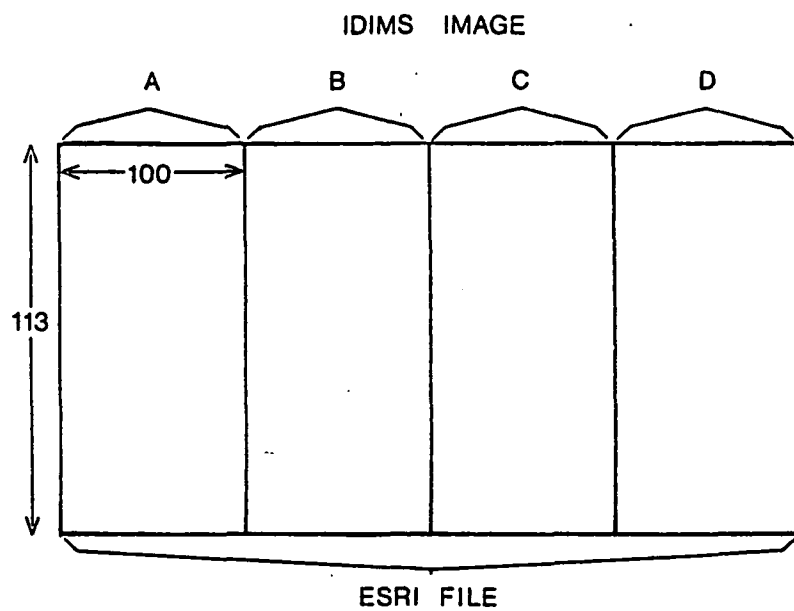
V. Conversion to Image Format

The next step involved transporting the grid data to NASA/ARC for use on the ESL Interactive Digital Image Manipulation System (IDIMS). The ESRI data sets were prepared, using a format believed to be compatible with IDIMS transfer function tapes for multiple four-band images. This method was not immediately successful. The tape was again processed at NASA/ARC, this time read as groups of four single-band images. This effort was also unsuccessful. Multiple record blocking, which should have solved the problem had not yet been implemented on the IDIMS system at NASA/ARC.⁴

A portion of this problem was traced to a spurious short record at the beginning of file one. By specifying single-band processing with labels for the first file, it was possible to read the data as 100 x 113 vertical strips from an implicit 400 x 113 image. (See Figure 5 for a graphical illustration of how the files were read as four strips of width 100 with starting line sample locations of (1,1) (1,101) (1,201), and (1,301).)

Succeeding files were similarly read, without label processing. Eventually, all the ESRI files were placed on the IDIMS transfer tapes for future operations. A more efficient format has been designed for future data transfers.

⁴ Problems encountered in this step are characteristic of the IDIMS and may or may not apply to all image processing systems.



ESRI Grid Cell Files as read by IDIMS

Figure 5

PROCEDURES IN PREPARATION AND CLASSIFICATION OF THE LANDSAT DATA

I. Obtain Landsat Data.

Personnel at NASA Ames Research Center obtained a raw Landsat data tape which was generated by Jet Propulsion Laboratory (JPL), using scenes from August 1976. The data had bulk geometric correction and had been spatially and spectrally mosaicked into 1° by 1° quadrangles. The data was then classified at NASA/ARC using unsupervised clustering on the ILIAC IV computer. This classification was done as part of the California Department of Forestry (CDF) project for the entire state with the state being divided into ecozones.⁵ The San Bernardino West quadrangle (containing Redlands) has three defined ecozones which are Desert, Los Angeles Basin, and South Border. For these ecozones there was a total of 167 clusters.

II. Define the Study Boundary

This step was performed by using the IDIMS and defining a window of the San Bernardino West quadrangle, which included the study area. The resulting window was a subset of the entire quadrangle and was 250 lines (rows) of 200 samples (columns). The parcel size was 80 meters nominal and map grid geometry was Lambert Conformal Conic with standard parallels at 33° and 45° , rotated such that grid north was 11° east of due north along 117° west longitude. Windows of both the classified and raw data (4 band) were created.

⁵ Specific issues and procedures used in this classification project are not addressed in this paper.

III. Registration of Data Set to UTM Grid.

A. Digitizing Control Points.

Within the study area, eleven control points were located. Using the IDIMS display curser command, the line-sample locations for each of these control points were displayed and recorded for later use. With the aid of an orthophoto quadrangle of the Redlands area, these control points were also located on the 7.5 minute quadrangle. Using the IDIMS geographic entry system (GES), the control points were digitized into a central geoblock for further use in IDIMS. The central points were typically street intersections, corners of cleared areas, intersections of airport runways, parking lots, and stream junctions. Image position estimates were done to the nearest whole pixel; actual locations were digitized on the map.

B. Creation of Transformation Records.

In this study, three transformation records were generated for registration purposes. Using IDIM's functions TRANSFORM and ALLCOORD utility programs, the GES control points were registered to the Landsat Control Points (T1), the GES control points were converted to UTM row and column equivalents and registered to the UTM grid (T2), and the UTM grid equivalents of the GES control points were registered to the Landsat control points (T3). Transformation matrix T1 is the most critical as it relates line sample values to latitude-longitude coordinates. For the registration process a third degree polynomial function was employed which, for the worst case, yielded a control point value that was one pixel width in error.

C. Geometric Correction and Registration.

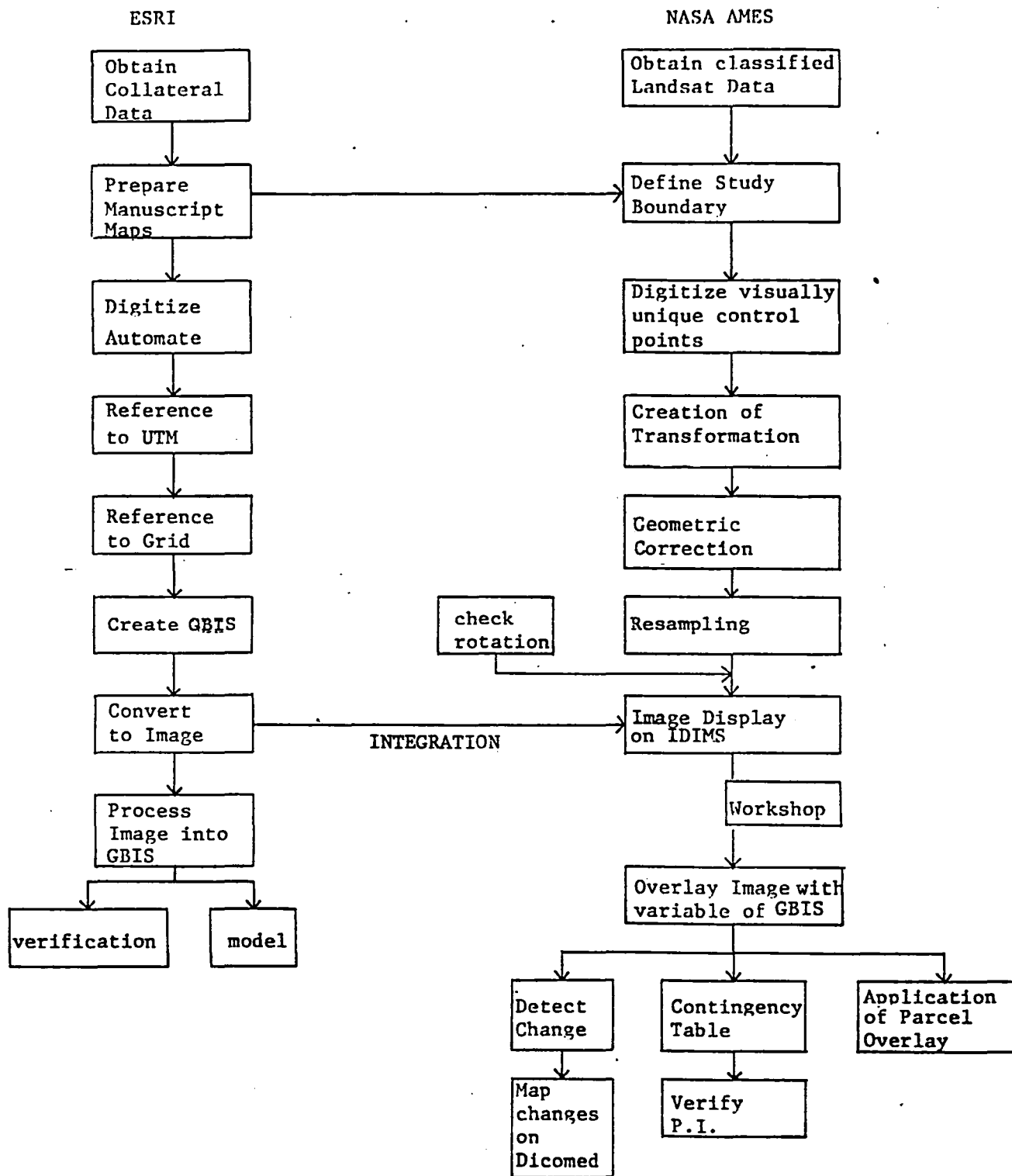
Using IDIMS function REGISTER, and the above transformation records, the window from the classified image was rotated and re-sampled using a nearest neighbor algorithm. After this procedure the study area (except three pixels) was contained within the Los Angeles Basin ecozone, and included 40 unsupervised clusters or classes. It contained 113 lines of 100 samples each with 80 meter pixels aligned with the UTM coordinate system.

IV. Conversion of Landsat Data to GBIS Format.

Personnel at NASA/ARC created a magnetic tape which was compatible with the ESRI computer facility. * Software was written which converted the Landsat data to a GBIS file format. There were no significant problems with this data transfer. This Landsat data was later integrated into the ESRI GBIS. This will be discussed in the next section.

FIGURE 6

FLOW CHART OF VERTICAL INTEGRATION PROCEDURES



INTEGRATION OF DATA SETS

Successful integration is eventually dependent upon the ability of the users to satisfactorily meet data needs. This achievement is dependent upon the complex interaction of a specific problem and local data characteristics of involved data systems. In a "perfect world", user data needs would shape data characteristics, and in turn determine system characteristics. In reality, data sets often either predate computer systems or are shaped by system capabilities. Fiscal and institutional constraints place additional limits on how the data sets may be integrated.

Three basic options for integration of data sets are:

1. Conversion of a geographic based information system (GBIS) file to Landsat image format, with subsequent processing in an image-processing environment.
2. Conversion of Landsat images to GBIS format, with subsequent processing on the GBIS.
3. Treatment of both a GBIS and Landsat data set in a hybrid system or through a least common denominator compromise.

Each option has its own set of problems as a result of conventional and Landsat data characteristics and the design limitations of the data system.

Only the first two of these options were explored in this pilot study. A flow chart of these procedures is illustrated in Figure 6.

I. Integration of GBIS Files with Landsat Data.

This project successfully integrated a grid cell data base with an image system format.⁶

Image systems, such as IDIMS, while being able to efficiently carry out image-related functions are often limited in their ability to manipulate external data contained in a geographic data base. With IDIMS, two options are available for external data entry:

1. Digitization of map materials using the geographic entry system which is a direct component of the image processing system. A possible solution to this option would be the conversion of "outside users" data to a format that the ESL digitizing equipment uses. But, internal data structures used by ESL, while enhancing system efficiency, are very difficult to replicate from externally generated data. This option would require considerable labor for data bases of any large content.
2. Conversion of GBIS files to pseudo-images. This was the option that was used for this integration effort. This option required that GBIS data files be converted to pseudo-images for use on IDIMS. As images are most effectively handled as byte data with values of 0-255, data values or code identifiers must be within this range. Nominal or labeled data sets must be converted to an index.

⁶ Procedures or problems encountered are related to the IDIMS and may or may not be representative of all image processing systems.

Image to image manipulation functions are extensive although their immediate application to geoprocessing problems are not clear. For example, some assessor's parcel data were digitized and automated with block number and parcel number identifiers. With IDIMS, the parcel data were overlaid with the Landsat data to determine the utility of this form of operation. The objective of this operation was to assign Landsat data values to the parcel. The usefulness of this type of data related to the parcel has obvious merits especially for county data systems. The resulting composite was only marginally useful, as parcel codes were not unique outside a particular block. Some conclusions can be made however. The coverage of the parcel data should have been full which implies that streets and other "non-parcel" data should be included in the data base. Parcel numbers or identifiers should be unique, even if this requires a certain redundancy in the data.

Also in the internal workshop at NASA/ARC, Landsat classified data was observed and compared with other geo-based files, such as 1974 land use. Contingency tables were generated, comparing both the land use file with classified Landsat data from the California Division of Forestry (CDF) project, which show the number of occurrences of codes between the two files and also calculating the expected value of these codes. Two contingency tables are included in this report:

- 1) Table One compares the 1974 land use file with the classified Landsat data.
- 2) Table Two compares the 1978 land use file with the classified Landsat data.

Table 1. 1974 Land Use vs. Classified Landsat Data

COLUMBIA AIRCRAFT OF 1915										COLUMBIA AIRCRAFT OF 1916									
COLUMBIA AIRCRAFT OF 1915										COLUMBIA AIRCRAFT OF 1916									
COLUMBIA AIRCRAFT OF 1915										COLUMBIA AIRCRAFT OF 1916									
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13
14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15
16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17	17
18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23
24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26
27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28
29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29

Table 2. 1978 Land Use vs. Classified
Landsat Data.

Using both of these tables, one could identify the changes in code frequencies between the files included in the contingency tables.

Each classified Landsat cluster was named for the major constituent class from the 1978 land use file. The 1974 land use file was also observed when change was apparent in the area. Comparing the 1974 and 1978 land use files was complicated somewhat by the fact that the minimum polygon resolution was smaller in the 1978 land use file than the 1974 file. If no single class was clearly dominant for the Landsat cluster, a compound name (i.e. orchard/brush) was used. From this study it was apparent that wide varieties of association existed between the geo-based files and the Landsat data. For example, one Landsat cluster labeled "brush" fell 94% into "orchard" in the land use file. At the other extreme, a cluster labeled "urban" in the classified file fell 86% into orchard of the land use file. This illustrates the problem of applying cluster labels developed for a large area (the Los Angeles Basin Ecozone) into some smaller constituent area. Table 3 illustrates a comparison of Landsat cluster values before and after the integration process and clearly demonstrates a finer breakdown of classification which resulted from the integration of the geo-based land use files with the Landsat data.

In another experiment, image "masks" were created, using a Boolean logic function. All 1974 land use pixels with agricultural uses were converted to a byte value of 127 and all other uses to a zero value. Using an ADD function, this mask was overlaid on the Landsat classified image. The result of this was an image that was background for all non-agricultural uses, showing Landsat classified values for agricultural areas only. The resultant image was used in a crude method of change detection between

Table 3

COMPARISON OF LANDSAT CLUSTER

VALUES BEFORE AND AFTER INTEGRATION PROCESS

<u>Landsat Cluster</u>	<u>CDF Group</u>	<u>Group Classification After Integration</u>
88	Grass	Grass/Golf Courses & Schools
90	Barren	Barren/Sand and Gravel
92	Barren	Barren/Sand and Gravel
93	Barren	Barren/Sand and Gravel
103	Grass	Grass/Golf Courses
106	Brush	Orchards
107	Brush	Steep Undeveloped
108	Brush	Orchards and Brush Including Abandoned
109	Brush	Orchards
110	Brush	Brush
111	Urban	Orchard/Transitional
112	Brush	Orchard
113	Brush	Brush/Orchard
114	Urban	Orchard/Residential
115	Urban	Urban/Orchard
		Residents/Orchard
116	Urban	Commercial
117	Urban	Residential
118	Urban	Residential/Mixed Urban
119	Urban	Commercial
120	Urban	Residential
121	Urban	Commercial
122	Grass	Residential/Transitional or Grass
123	Grass	Pavement
124	Grass	Pavement

1974 and 1976. However, there were a large number of "false positives" due to the smaller effective minimum mapping unit employed in the Landsat classified image.

Using another IDIMS function, polaroid positives as well as reproducible negatives were created using a DICOMED film recorder for the 1974 land use, 1976 Landsat classes, and the 1976 Landsat masked only for agricultural areas. Comparable, but not identical, color schemes were used to facilitate visual image comparison. A notable feature is the visual comparison problem caused by the relatively fine texture of the Landsat data. Reproductions of these films have been done, but are not included in this report.

As a final experiment, terrain data derived by JPL from DMA/NCIC digital terrain files were also registered to the study area grid. As the terrain data was already aligned with the Landsat data by JPL, no additional control points were needed and the existing transformation equivalencies were applied. Several problems were seen in the resultant images. Nearest neighbor resampling is probably not appropriate for terrain data. As an example, registered terrain data did not yield the same slope value for many pixels, due to the nearest-neighbor resampling method changing the values of adjacent pixels used for slope calculations. Similarly, nearest neighbor resampling distorted the JPL aspect data for the same reason. Additionally, the aspect data from JPL was oriented toward Lambert grid north, not true north. When resampled, it was still relative to Lambert grid north, and therefore of little apparent value in a UTM geo-based file or the registered and resampled Landsat file. Aspect data pointed from true north would have resampled correctly.

Performing data integration in the IDIMS environment has major advantages and disadvantages. Landsat unsupervised cluster labelling is greatly facilitated by immediate, interactive access of registered collateral files. The interactive display allows for rapid evaluation of various cluster grouping and mapping strategies to maximize final image/product information content. Processing times fast, and allow for rapid evaluation of various modeling constructs and sensitivity analyses in both quantitative and qualitative frameworks. Unfortunately, IDIMS' ability to perform complex modeling is severely constrained by a cumbersome method of Boolean relational modeling; while an experienced user might be able to successfully concentrate various image-manipulation functions required for such modeling, the occasional user would likely be overwhelmed. Another serious problem is the inability of the system to accept external polygon files as input or produce polygon files as output for other systems. IDIMS is also largely incapable of handling external files in the case where image data values may simply serve as pointers to such a file, i.e., an image of soil series names/codes that reference an extensive series-specific table of soil characteristics.

II. Integration of Landsat Classified Image with a GBIS.

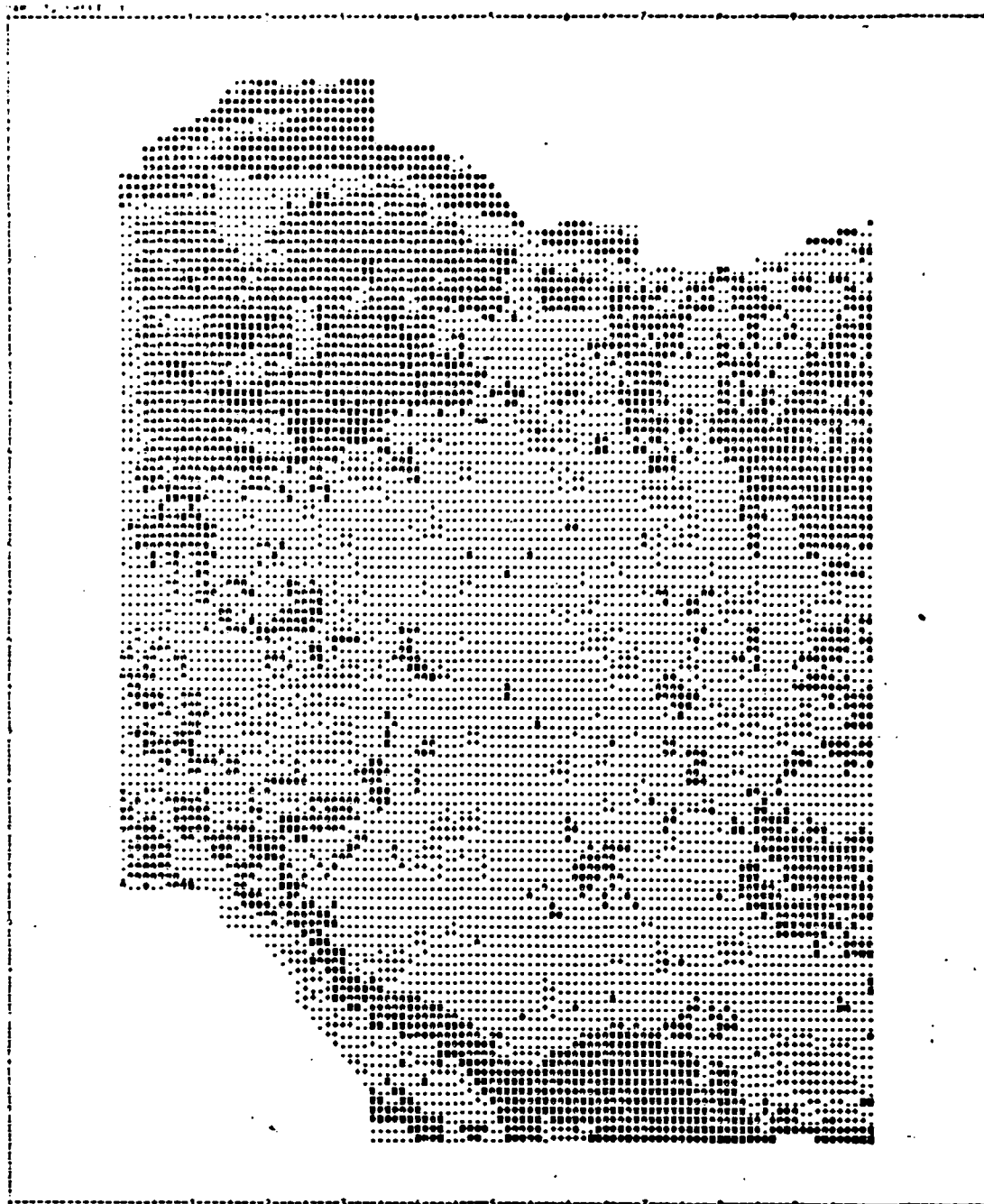
Personnel at ESRI also successfully integrated the Landsat image into a GBIS data base. Initially it is seen that the Landsat data could serve two functions. First, the Landsat data could be used as collateral data to verify the photo interpretation used in establishing the GBIS data base. Secondly, the Landsat data, once accurately classified, could be used in the modeling processes of land capability and land suitability studies. At the time of this publication, the first of these options had not fully been explored, but considerable work has been done with using the Landsat data for grid cell modeling. For example, the Landsat data was used in the GBIS data base to generate grid cell area calculations for various types of land use between 1974, 1976, and 1978. Figure 7 displays a grid cell map that was generated with existing ESRI software of the Landsat data.

Although image systems are generally similar in their treatment of data sets, GBIS vary considerably not only in their internal data representation, but also in the way the data sets are manipulated. Several basic data base concepts are in common use:

1. Regular grid cell, either square or rectangular (ESRI data).
2. Triangular or other non-rectangular grid cell.
3. Polygon systems (either closed loop or intersection-chain networks).
4. Mixed systems.

Each system will have its own set of technical problems when dealing with integration. The following subset of problems will be common to nearly all systems:

A. Effective Spatial Resolution: Most GBIS data bases employ a data configuration other than an 80 meter square pixel. A mechanism (such as



DATA AMP, WENT STAY

REDUCTION ANGLE

LANDSAT DATA

CLASSIFICATIONS	
1	WATER
2	FOREST
3	AGRICULTURE
4	URBAN

DATA HADEN IN 5 LEVELS BETWEEN POTENTIAL VALUES OF 1 AND 5

ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

MINIMUM

PERCENTAGE OF TOTAL ABSOLUTE VALUE RANGE APPLYING TO EACH LEVEL

FREQUENCY DISTRIBUTION OF DATA POINT VALUES IN EACH LEVEL

LEVEL

STAY

POTENTIAL

Figure 7 - Classified Landsat Data

in IDIMS) for resampling, aggregating, or other routines is necessary to assign classified Landsat values to GBIS geographic units. Users will often say that Landsat's nominal 80 meter resolution is too coarse for their applications; yet classified image products may be criticized as being too detailed. This effect is related to the more discipline-related problem of spectral versus spatial classification and resolution. One solution might be in an algorithm designed to "group" classified data into geographically contiguous areas to simulate the minimum mapping unit criteria applied by photo interpreters.

B. Geographic Referencing: Two issues are important:

1. Projection: GBIS will generally use either state plane grid coordinates, universal transverse mercator (UTM) grid coordinates, or latitude-longitude coordinates. Geometrically corrected Landsat data will be in whatever projection was used to generate the transformation equation. In the CDF project a Lambert Conformal Conic projection was used. While this projection may be useful for a statewide mosaic, it is not useful for small areas. Projection changes may be accomplished utilizing either the geometric properties of the projection or a polynomial transformation, depending upon the situation.
2. Geometric Precision: When using Landsat data with other data sets, it is extremely important to know the geographic location of any given pixel. The degree of precision to which this can be accomplished has both practical (time, cost, etc.) as well as theoretical (integrating scanner) limitations.

C. System-Specific Considerations (First Impressions).

1. Regular Grid Cell: If grid cell orientation and projections are the same, integration is simple. However, misalignment problems would be very difficult to solve in a grid system. Software for performing grid data base rotation and resampling is not included in most grid-oriented systems. Solutions range from image rotation and rescaling as in image-based systems to manual grid cell encoding from hard copy images.
2. Non-Rectangular Grid Cell: This would likely require an even more complex interpolation routine. Other considerations are similar to the regular grid cell discussion above.
3. Polygon Systems (closed loop): Integration would likely require a point-in-polygon interpolation routine. Landsat data might have to be converted to polygon data sets, a non-trivial operation. Manual re-digitizing might be useful as a technique, but quite cumbersome.
4. Polygon Systems (Intersection-Chain Networks): This is probably the most difficult interface without some system hybridization and shares the same problems with closed loop.
5. Mixed Systems: A number of polygon systems are in fact mixed systems, with integrated polygon-to-grid conversion routines and grid cell manipulation capabilities. Such systems would seem to offer the greatest interfacing potential with a minimum of new software design. (The ESRI software package is a good example; there are others, including the USGS Geographic Information Retrieval and Analysis System currently under development.)

Performing Landsat integration in the traditional GBIS environment also has its strengths and weaknesses. Processing speeds are slower compared to image systems, as most grid-based systems are not designed to manipulate very large data files.⁷ They are designed to manipulate sample data files which is typical of Landsat. Modeling capabilities in conventional systems are generally more user-oriented, and display and model analysis rarely occur in an interactive mode. Traditional outputs include printer maps, electrostatics, and DICOMED images.

Landsat's relatively small cell size will often require aggregation when integration occurs. Hence, informational detail will be lost in systems using much larger, e.g. 40 acre grid cells. Another potential problem in the integration process is that most GBIS are incapable of performing precise registration between two grid data sets with dissimilar map projections, cell sizes, and grid orientations.

7 These data files are considered to be 1000 x 1000 or greater.

CONCLUSION

Resource planning/management responsibilities and activities at all levels of government have greatly increased the demand and use of environmental data. Governments are increasingly turning to computerized geo-based information systems to handle these large volumes of data. Dealing with the costs of these data sets and data systems is becoming increasingly difficult as governments face rising costs and limited revenues.

One potential answer to these problems involves vertical data integration, defined as the creation of data sets and accompanying analysis techniques that may be used by various levels of government. Vertical data integration should allow cost-sharing of data sets and analysis techniques, minimum duplication, and especially with high-technology data sources like Landsat, considerable potential for realizing scale economies and improved technology transfer.

Based on our experiences of these two integration efforts, the evaluation of any Landsat-GBIS integration process requires:

1. Investigation of the conventional data sources.
2. Investigation of the characteristics of Landsat data products.
3. Characteristics of data systems and analysis techniques used in the integration process.

Landsat data has some known advantages as a large area data source, such as

- 1) digital multi-band format which allows for low-cost spectral classification;
- 2) uniformity within the data set and compatibility between data sets; and
- 3) timelines and currency which can provide a significant temporal components.

These advantages are offset by problems relating to the Landsat product itself - such as format, registration/geometric correction, image quality and resolution,

as well as with classifications of the digital image which may be project or discipline-specific. Landsat may be interfaced with conventional data through a number of operating modes, including: 1) manual photo interpretation with subsequent manual or semi-automatic entry of resulting graphic data into a GBIS; 2) utilization of digital products in image-based system; or 3) utilization of digital products or derived data bases with non-image systems. Evaluation of these modes of data integration requires the investigation of several geographic data handling problems, such as polygon-grid conversion, scale and projection change, image manipulation, and hybrid systems, within the context of data system limitations.

These issues will be investigated within the context of actual county government operations in California with emphasis placed on integrating Landsat derived data into existing data bases used to analyze environmental problems. County governments represent a major potential user community with a pressing requirement for environmental and resources data; additionally, their GBIS activities may be taken as representative of similar activities undertaken by both regional and local governments.

Evaluation of any proposal for the vertical integration of Landsat data into the California governmental or private usage pattern requires addressing and resolving a number of technical and institutional issues which include:

- a) The degree of centralization of digital image preprocessing activities.
- b) The degree of centralization of interpretation (digital analysis) activities.
- c) An evaluation of the utility of various product types at various levels of user sophistication.
- d) Determination of the flexibility versus efficiency trade-offs in generating digital products for use at various governmental levels.
- e) Determination of the limiting effects of spatial resolution and geo-

referencing of Landsat data sets and near-future satellite sensors in urban areas with particular application to change detection potential.

- f) Evaluation of the feasibility of a centralized, stage government operated Landsat processing utility function.

A test site of San Bernardino County has been selected where county, state, and federal agencies will participate in a NASA sponsored project of vertical integration to study these issues.

The objectives of the project may be grouped as follows:

- a) Describe and evaluate the parameters of conventional data sources, Landsat products, and both operational and near-operational GBIS systems that influence the technical feasibility of Landsat integration. Such analysis will likely concentrate on issues relating to geographic referencing and classification techniques.
- b) Investigate the data requirements of environmental planning and resource management activities of county government with emphasis on those disciplinary areas where Landsat has the greatest application potential.
- c) Select a set of test problems with emphasis on development and demonstration of techniques for integrating Landsat data into locally operated GBIS. First year work will evaluate the utility of classified Landsat digital data provided by higher levels of government.
- d) Demonstrate project findings through the generation of agency products from GBIS systems utilizing integrated Landsat data.
- e) Document the development of data parameters, system descriptions, application selection, and test results, including costs and benefits.
- f) Describe implications of test results as potential for future technology transfer, related applications, and future data system development.

1. Report No. NASA CR-166352	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Exploration into Technical Procedures for Vertical Integration		5. Report Date September 1979	
		6. Performing Organization Code	
7. Author(s) Russell J. Michel Keith D. Maw		8. Performing Organization Report No.	
9. Performing Organization Name and Address Environmental System Research Institute 380 New York St. Redlands, CA 94035		10. Work Unit No. T3869	
		11. Contract or Grant No. A-65111B p.c.	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		13. Type of Report and Period Covered Final Contractor Report	
		14. Sponsoring Agency Code	
15. Supplementary Notes Technical Monitor - Susan Norman, M/S 242-4 NASA Ames Research Center TE: 415-965-5912 Moffett Field, CA 94035 FTS 448-5912			
16. Abstract This project explored issues in the design and use of a digital geographic information system (GIS) incorporation landuse, zoning, hazard, Landsat, and other data. In the project an eleven layer data base was generated for the Redland, California 7½" quadrangle. Examined were issues in spatial resolution, registration, grid versus polygonal structures, and comparison of photointerpreted landuse to Landsat land cover. This work has been used as the basis for designing a number of projects to be carried out in FY80 and 81 as part of the California Integrated Remote Sensing System (CIRSS) ASVT.			
17. Key Words (Suggested by Author(s)) Landsat GIS Geographic Information System CIRSS		18. Distribution Statement Unclassified-Unlimited STAR Category 43	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 30	22. Price*

End of Document